

# Cross-Scale Predictive Dictionaries

Vishwanath Saragadam  
vishwanathsrv@cmu.edu

Aswin Sankaranarayanan  
saswin@andrew.cmu.edu

Xin Li  
xinli@cmu.edu

Dept. of ECE, Carnegie Mellon University

## Abstract

We propose a novel signal model, based on sparse representations, that captures cross-scale features for visual signals. We show that cross-scale predictive model enables faster solutions to sparse approximation problems. This is achieved by first solving the sparse approximation problem for the downsampled signal and using the support of the solution to constrain the support at the original resolution. The speedups obtained are especially compelling for high-dimensional signals that require large dictionaries to provide precise sparse approximations. We demonstrate speedups in the order of  $10 - 100\times$  for denoising and up to  $15\times$  speedups for compressive sensing of images, videos, hyperspectral images and light-field images.

## 1. Introduction

Visual signals exhibit strong correlation across scales that can often be modeled and exploited to enhance image processing algorithms [2, 27]. An important example of this idea is the multi-scale coding of images using the wavelet-tree model which provides both a sparse as well as a predictive model for the occurrence of non-zero wavelet coefficients across scales [32]. Specifically, the wavelet tree model arranges the wavelet coefficients of an image onto a tree such that nodes on the tree correspond to the coefficients and each level corresponds to coefficients associated with a particular scale. Under such an organization, the dominant non-zero coefficients form a connected rooted sub-tree[5], i.e., children of a node with small wavelet coefficients are expected to take small values as well. The wavelet tree model is central to many compression [28], sensing [7, 11], and processing algorithms [5].

Learnt dictionaries provide an alternate approach to wavelets in terms of enabling sparse representations [24]. Given a large amount of data, there are many approaches that learn a dictionary such that the training dataset can be expressed as a sparse linear combination of the elements/atoms of the dictionary. The reliance on machine



Figure 1: Left to right: Bayer image, image reconstructed using OMP, and image reconstructed using our proposed method. While OMP takes 16 minutes, our proposed method takes only 1.5 minutes.

learning, as opposed to analytic constructions as in the case of wavelets, provides immense flexibility towards obtaining a dictionary that is tuned to the specifics of a particular signal class or application. Yet, in spite of a large body of work devoted to learning sparse representations, there is little work devoted to learning predictive models that exploit correlations across spatial, temporal, spectral and angular scales. In this paper, we propose a multi-scale dictionary model for visual signals that naturally enables cross-scale prediction. Our contributions are as follows.

- **Model.** We propose a novel signal model that uses multi-scale sparsifying dictionaries to provide cross-scale prediction for a wide array of visual signals. Specifically, given the set of sparsifying dictionaries — one for each scale — the non-zero support patterns of a signal and its downsampled counterparts are constrained to only exhibit specific pre-determined patterns.
- **Computational speedups.** We show that the proposed signal model, with its constrained support pattern across scales, naturally enables cross-scale prediction that can be used to speedup algorithms like OMP. We term our algorithm, *zero tree OMP*, as the sparse representation of the signal forms a zero tree under the current model.
- **Learning.** Given large collections of training data, we propose a simple training method, which is a modified form of K-SVD training method, to obtain dictionaries that are consistent with our proposed model.

- **Validation.** We verify empirically that the model works through simulation on an array of visual signals including images, videos, hyper-spectral and light-field images.

The organization of this paper is as follows. In Section 2, we present related work in the area of sparse representation and dictionary learning. Section 3 introduces the proposed model, details the benefits of the model in terms of speedup, and finally, presents an approach to learn the model given training data. In Section 4, we validate our approach on a range of visual signals to verify our model.

## 2. Prior work

**Notation.** We denote vectors in bold font and scalars/matrices in capital letters. A vector is said to be  $K$ -sparse if it has at most  $K$  non-zero entries. The list of indices of non-zero entries of a sparse vector is termed its support; the support of a vector  $\mathbf{s}$  is denoted as  $\Omega_{\mathbf{s}}$ . The  $\ell_0$ -norm of a vector is the number of non-zero entries. Finally, given a dictionary  $D \in \mathbb{R}^{N \times T}$  and a support set  $\Omega$ ,  $D|_{\Omega}$  refers to the matrix of size  $N \times |\Omega|$  formed by selecting columns of  $D$  corresponding to the elements of  $\Omega$ ; similarly, given a vector  $\mathbf{s}$ ,  $\mathbf{s}|_{\Omega}$  refers to an  $|\Omega|$ -dimensional vector formed by selecting entries in  $\mathbf{s}$  corresponding to  $\Omega$ .

**Sparse approximation.** Sparse approximation problems arise in a wide range of settings [12]. The broad problem definition is as follows: given a vector  $\mathbf{x} \in \mathbb{R}^N$ , a matrix  $D \in \mathbb{R}^{N \times T}$ , we solve

$$(P0) \quad \min_{\mathbf{s} \in \mathbb{R}^T} \|\mathbf{x} - D\mathbf{s}\|_2 \quad \text{s.t.} \quad \|\mathbf{s}\|_0 \leq K.$$

There are many approaches to solving (P0) and its many variants. Of particular interest to this paper is orthogonal matching pursuit (OMP) [25], a greedy approach to solving (P0). OMP recovers the support of the sparse vector  $\mathbf{s}$ , one element at a time, by finding the column of the dictionary that is most correlated with the current residue. In each iteration of the algorithm, there are three steps: first, the index of the atom that is closest in angle to the current residue is added to the support; second, solving a least square problem with the updated support to obtain the current estimate; and third, updating the residue by removing the contribution of the current estimate.

**Speeding up OMP.** Obtaining sparse representations with high accuracy often requires dictionaries with a large number of redundant elements. Figure 2 shows the timing vs. accuracy for a dictionary of  $8 \times 8$  image patches for varying number of dictionary atoms,  $T$ . We observe that the increase in accuracy enabled by a dictionary with larger number of atoms comes with increased computational time as

well. A number of techniques have been devoted to speeding up different aspects of the problem. For problems in high-dimensionality, i.e. large  $N$ , one approach is to embed to work on random projections of the dictionary [31]. Specifically, as opposed to the objective  $\|\mathbf{x} - D\mathbf{s}\|_2$ , we minimize  $\|\Phi\mathbf{x} - \Phi D\mathbf{s}\|_2$  where  $\Phi \in \mathbb{R}^{M \times N}$ ,  $M < N$ , is a random matrix that preserves the geometry of the problem thereby allowing us to do all computations in an  $M$ -dimensional space. In the context of high-dimensional data, it is typical to have dictionaries with a very large number of atoms [16], i.e.,  $T \gg N$ . Here, the search for the atom closest to the residue becomes the most time-consuming step. One approach to speeding up OMP is by using approximate nearest neighbors and shallow-tree based matching [4, 15]. Another approach is to restrict the search space by imposing a tree structure on sparse coefficients [18]. Speed up in OMP has also been obtained through parallel implementation of the search for atoms [13], and through tweaking the least squares step [14]. However, such methods provide lesser improvements for very large problem sizes.

**Dictionary learning.** For signal classes that have no obvious sparsifying transforms, a promising approach is to *learn* a dictionary that provides sparse representations for the specific class of interest. Field and Olshausen [24], in a seminal contribution, showed that patches of natural images were sparsified by a dictionary containing Gabor-like atoms — this provided a connection between sparse coding and the receptor fields in the visual cortex. More recently, Aharon et al. [3] proposed the “K-SVD” algorithm which can be viewed as an extension of the k-means clustering algorithm for dictionary learning. Given a collection of training data  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_Q]$ ,  $\mathbf{x}_i \in \mathbb{R}^N$ , K-SVD aims to learn a dictionary  $D \in \mathbb{R}^{N \times T}$  such that  $X \approx D[\mathbf{s}_1, \dots, \mathbf{s}_Q]$  with each  $\mathbf{s}_k$  being  $K$ -sparse.

**Multi-scale dictionary models.** The idea of coupling multi-scale models and sparsifying dictionaries has been explored before. Jayaraman et al. [30] provide a multi-level representation of image patches where simple patches with little textures are captured in the early stages while more complex textures are only resolved at the higher levels. This provides speedups when solving sparse approximation problems since patches that occur more often are captured at the earlier levels. Jenatton et al. [17] present a hierarchical dictionary learning mechanism, where they impose a tree structure on the sparsity, which forces the dictionary atoms to cluster like a tree. Though it does give higher accuracy of reconstruction, not much has been said about speed up obtained. Mairal et al. [22] learn a dictionary based on quad-tree models, where each patch is further sub-divided into four non-overlapping patches. While this method gives better accuracy, the algorithm is very slow, as claimed by

the authors. None of the multi-scale learning algorithms exploit the cross-scale coding especially for visual signals.

The goal of this paper is to construct dictionaries endowed with structured sparse representations, similar to the wavelet-tree model, and enable computational speedups in solving sparse approximation problems.

**Compressive sensing (CS).** An application of sparse representations is in CS where we sense signals from far-fewer measurements than their dimensionality [6]. CS relies on the low-dimensional representations for the sensed signal — sparse representation under a transform or a dictionary being an example of this. There is a rich body of work on applying compressive sensing to imaging or sensing visual signals including videos [16], light fields [23, 29], and hyperspectral images [8, 19, 20]. Most relevant to our paper is the video CS work of Hitomi et al. [16] where a sparsifying dictionary is used on video patches to recover high-speed videos from low-frame rate sensors. Hitomi et al. also demonstrated the accuracy enabled by very large dictionaries; specifically, they obtain remarkable results with a dictionary of  $T = 100k$  atoms for video patches of dimension  $N = 7 \times 7 \times 36 = 1764$ .

**Wavelet-tree model.** Our proposed method is inspired by multi-resolution representations and tree-models enabled by wavelets. In particular, Baraniuk [5] shows that the non-zero wavelet coefficients form a rooted sub-tree for signals that have trends (smooth variations) and anomalies (edges and discontinuities). Hence, piecewise-smooth signals enjoy a sparse representation with a structured support pattern with the non-zero wavelet coefficients forming a rooted sub-tree. Similar properties have also been shown for 2D images under the separable Haar basis [28]. However, in spite of these elegant results for images, there are no obvious sparsifying bases for higher-dimensional visual signals like videos and light-field images. To address this, we build cross-scale predictive models, similar to the wavelet tree model, by replacing a basis with an over-complete dictionary that is capable of providing sparse representation for a wide class of signals.

### 3. Proposed signal model

**Proposed cross-scale predictive sparse model.** We propose a signal model that predicts the support of a signal across scales (see Figure 3). For simplicity, we first present the model for a two-scale scenario.

Given a collection of signals,  $\mathcal{X} \subset \mathbb{R}^N$ , our proposed signal model consists of two sparsifying dictionaries  $D_{\text{high}} \in \mathbb{R}^{N \times T_{\text{high}}}$  and  $D_{\text{low}} \in \mathbb{R}^{N_{\text{low}} \times T_{\text{low}}}$  that satisfy the following three properties.

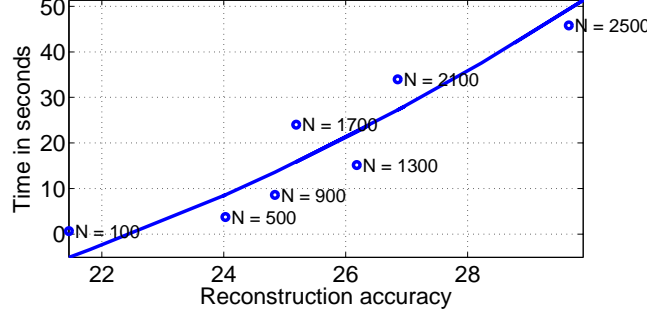


Figure 2: Plot of approximation time versus approximation accuracy for varying dictionary size for  $8 \times 8$  image patches.

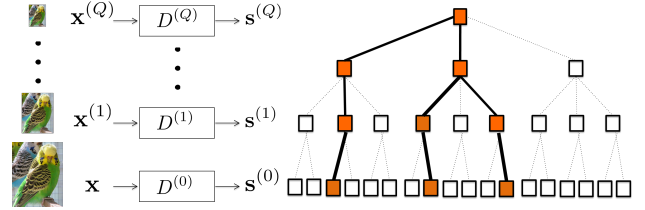


Figure 3: Proposed cross-scale signal model with sparse coefficients across scales forming a rooted subtree. A child can be nonzero only if the parent is nonzero.

- *Sparse approximation at the finer scale.* A signal  $\mathbf{x} \in \mathcal{X}$  enjoys a  $K_{\text{high}}$ -sparse representation in  $D_{\text{high}}$ , i.e.,  $\mathbf{x} \approx D_{\text{high}} \mathbf{s}_{\text{high}}$  with  $\|\mathbf{s}_{\text{high}}\|_0 \leq K_{\text{high}}$ .
- *Sparse approximation at the coarser scale.* Given  $\mathbf{x} \in \mathcal{X}$  and a downsampling operator  $W : \mathbb{R}^N \mapsto \mathbb{R}^{N_{\text{low}}}$ , the downsampled signal  $\mathbf{x}_{\text{low}} = W\mathbf{x}$  enjoys a sparse representation in  $D_{\text{low}}$ , i.e.,  $\mathbf{x}_{\text{low}} \approx D_{\text{low}} \mathbf{s}_{\text{low}}$  with  $\|\mathbf{s}_{\text{low}}\|_0 \leq K_{\text{low}}$ . The downsampling operator  $W$  is domain specific.
- *Cross-scale prediction.* The support of  $\mathbf{s}_{\text{high}}$  is constrained by the support of  $\mathbf{s}_{\text{low}}$ ; specifically,  $\Omega_{\mathbf{s}_{\text{high}}} \subset f(\Omega_{\mathbf{s}_{\text{low}}})$ , where the mapping  $f(\cdot)$  is known a priori.

We make a few observations.

*Observation 1.*  $T_{\text{high}} \gg T_{\text{low}}$  since  $N_{\text{high}} \gg N_{\text{low}}$ . With the increase of dimension of the signal, more complex patterns emerge which require larger number of redundant elements. Empirically we found that the number of atoms in a dictionary increases super linearly with increasing dimension of the signal for a given approximation accuracy.

*Observation 2.* Recall that the computational time for OMP is proportional to the number of atoms in the dictionary since, at each iteration of the algorithm, we need to compute the inner product between the residue and the atoms in the dictionary. If we can constrain the search space by constraining the number of atoms, then we can obtain computational speedups.

The proposed model obtains speedups by first solving a sparse approximation problem at the coarse scale and subsequently exploiting the cross-scale prediction property to constrain the support at the finer scale. The source of the speedups relies on two intuitive ideas: first, solving a sparse approximation problem for a problem with fewer atoms (and in smaller dimensions) is faster due to OMP's runtime being linear in the number of atoms of the dictionary used[21]; and second, if we knew the support of  $\mathbf{s}_{\text{low}}$ , then we can simply discard all atoms in  $D_{\text{high}}$  that do not belong to  $f(\Omega_{\mathbf{s}_{\text{low}}})$  since the support of  $\mathbf{s}_{\text{high}}$  is guaranteed to lie within  $f(\Omega_{\mathbf{s}_{\text{low}}})$ .

**Cross-scale mapping.** We use a simple strategy for the cross-scale mapping  $f$ . Let  $Q = T_{\text{high}}/T_{\text{low}}$  (assuming  $T_{\text{high}}$  and  $T_{\text{low}}$  are chosen to ensure  $Q$  is an integer). The cross-scale prediction map is defined using this simple rule.

$$i \in \Omega_{\mathbf{s}_{\text{low}}} \implies (i-1)Q + \{1, 2, \dots, Q\} \subset f(\Omega_{\mathbf{s}_{\text{low}}})$$

Each element of the support  $\Omega_{\mathbf{s}_{\text{low}}}$  in the coarser scale controls the inclusion/exclusion of a *non-overlapping block* of locations for the sparse vector in the finer scale. As a consequence, the cardinality of  $f(\Omega_{\mathbf{s}_{\text{low}}})$  is simply  $QK_{\text{low}}$ .

**Solving inverse problems under the proposed signal model.** We now detail the procedure for solving a sparse approximation problem using the proposed signal model (see Figure 3). Specifically, we seek to recover  $\mathbf{x} \in \mathcal{X}$  from a set of linear measurements  $\mathbf{y} \in \mathbb{R}^M$  of the form

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{e} = \Phi D_{\text{high}} \mathbf{s}_{\text{high}} + \mathbf{e},$$

where  $\Phi \in \mathbb{R}^{M \times N}$  is the measurement matrix and  $\mathbf{e}$  is the measurement noise. As indicated earlier, we obtain  $\mathbf{s}_{\text{high}}$  using a two-step procedure.

*Step 1 — Sparse approximation at the coarse scale.* We first solve the following sparse approximation problem:

$$(P_{\text{low}}) \quad \hat{\mathbf{s}}_{\text{low}} = \arg \min_{\mathbf{s}_{\text{low}}} \|\mathbf{y} - \Phi U D_{\text{low}} \mathbf{s}_{\text{low}}\|_2$$

$$\text{s.t.} \quad \|\mathbf{s}_{\text{low}}\|_0 \leq K_{\text{low}}.$$

Here,  $U : \mathbb{R}^M \mapsto \mathbb{R}^N$  is an upsampling operator such that  $WU$  is an identity map on  $\mathbb{R}^{N_{\text{low}}}$ . In all our experiments, we used a uniform down sampler and a nearest neighbour up sampler specific to the domain of the signal.

This first step recovers a low-resolution approximation to the signal,  $\mathbf{x}_{\text{low}} = D_{\text{low}} \hat{\mathbf{s}}_{\text{low}}$ .

*Step 2 — Sparse approximation at the finer scale.* Armed with the support  $\hat{\Omega} = \Omega_{\hat{\mathbf{s}}_{\text{low}}}$ , we can solve for  $\mathbf{s}_{\text{high}}$  by solving:

$$(P_{\text{high}}) \quad (\hat{\mathbf{s}}_{\text{high}})_{|f(\hat{\Omega})} = \arg \min_{\alpha} \|\mathbf{y} - \Phi(D_{\text{high}})_{|\hat{\Omega}} \alpha\|_2$$

$$\text{s.t.} \quad \|\alpha\|_0 \leq K_{\text{high}}.$$

The sparse approximation problems in both steps are solved using OMP. The proposed mapping across scales for the sparse support forms a zero tree, where a coefficient is zero if the corresponding coefficient at coarser scale is zero. Hence we refer to our algorithm as zero tree OMP.

**Theoretical speedup.** We next provide precise expressions for the expected speedups over the traditional single-scale OMP. Since any analysis of speedup has to account for the complexity of implementing  $\Phi$ , we consider the denoising problem where  $\Phi$  is the identity map.

Let  $C(N, T, K)$  be the amount of time required to solve a sparse approximation problem using OMP for a dictionary of size  $N \times T$  and sparsity level  $K$ . Hence, obtaining  $\mathbf{s}_{\text{high}}$  directly from  $\mathbf{x}$  would require  $C(N, T_{\text{high}}, K_{\text{high}})$  computations. In contrast, our proposed two-step solution using cross-scale prediction has a computational cost of  $C(N_{\text{high}}, T_{\text{low}}, K_{\text{low}}) + C(N, QK_{\text{low}}, K_{\text{high}})$ .

To compute the dependence of  $C(N, T, K)$  on  $N, T$ , and  $K$ , recall that for each iteration in the OMP algorithm, we need  $O(NT)$  operations [21] for finding inner product between the residue and the dictionary atoms,  $O(T)$  operations to find the maximally aligned vector and  $O(K^3 + K^2N)$  operations for the least-squares step. Thus,

$$C(N, T, K) = O(NTK + TK + K^4 + K^3N).$$

For dictionaries with a large number of atoms, i.e., large  $T$ , and small values for sparsity level  $K$ , the linear dependence on  $N$  dominates the total computation time. Here, the speedup provided by our algorithm is approximately  $T_{\text{high}}/(T_{\text{low}} + K_{\text{low}}Q)$ .

**Learning cross-scale sparse models.** We can learn the dictionaries  $(D_{\text{high}}, D_{\text{low}})$  with a simple modification to the K-SVD algorithm.

*Inputs.* The inputs to the learning/training phases are the training dataset  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$  and the values for the parameters  $K_{\text{high}}, K_{\text{low}}, T_{\text{high}}$ , and  $T_{\text{low}}$ .

*Step 1 — Learning  $D_{\text{low}}$ .* We learn the coarse-scale dictionary  $D_{\text{low}}$  by applying K-SVD to downsampled training dataset  $X_{\text{low}} = [W\mathbf{x}_1, W\mathbf{x}_2, \dots, W\mathbf{x}_n]$ . As a by-product of learning the dictionary are the supports  $\{\Omega_{\mathbf{s}_{\text{low}}, k_{\text{low}}}\}$  of the sparse approximations of the downsampled training dataset.

*Step 2 — Learning  $D_{\text{high}}$ .* We learn the fine-scale dictionary  $D_{\text{high}} = [\mathbf{d}_1, \dots, \mathbf{d}_{T_{\text{high}}}]$  by solving

$$\min_{D_{\text{high}}, S_{\text{high}}} \|Y - D_{\text{high}} S_{\text{high}}\|_F \text{ s.t. } \|\mathbf{d}_k\|_2 = 1,$$

$$\text{support}(\mathbf{s}_{\text{high}}) \subset f(\Omega_{\text{low}, k})$$

The above optimization problem can be solved simply by modifying the sparse approximation step of K-SVD to restrict the support appropriately. Figure 4 shows an example of the learned low resolution atoms and the corresponding



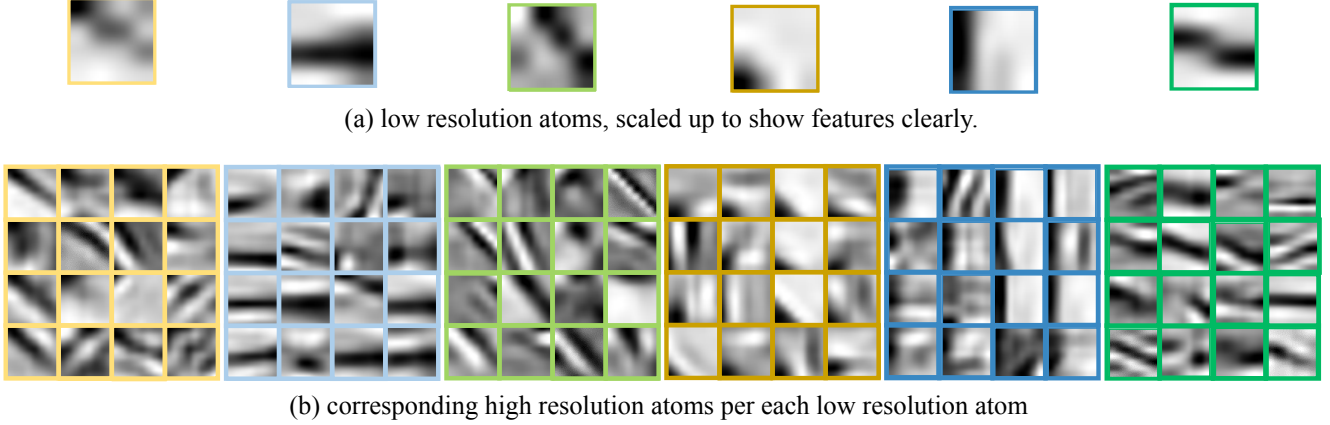


Figure 4: Visualization of select low-resolution atoms and their corresponding atoms in the high resolution dictionary.

high resolution atoms. Observe that constraining the sparse support of the high resolution approximation alone learns patches which are very similar in appearance to the low resolution patches, which is in strong favor of our signal model. We also note that the time required for training the model was about the same as that required to learn a single high-resolution dictionary, with the same specification as the high-resolution dictionary, using K-SVD.

**Parameter selection.** The design parameters in the two scale dictionary training are  $K_{low}$ ,  $K_{high}$ ,  $T_{low}$ , and  $Q$ .  $K_{low}$  can be chosen to fine tune the accuracy at lower scales. We found that  $K_{low} = 6$  to 10 provided high model approximation accuracy.  $K_{high}$  must be at least  $K_{low}$ , since at least one patch corresponding to each low resolution patch will be combined in high resolution.  $K_{high}$  can be increased further to increase approximation accuracy without much reduction in speed up. The value of  $Q$ , number of degrees of freedom for super resolving the low resolution atom, is highly dependent on the ratio  $N/N_{low}$ . For our experiments,  $N/N_{low} \approx 16$ , and hence  $Q$  was chosen as 16 as well.

#### 4. Experimental results

We compare zero tree OMP using our proposed two-scale dictionaries against traditional OMP on dictionaries learnt using K-SVD. We compare both the run time and approximation accuracy for images, videos, hyperspectral images and light-field images. We quantify approximation accuracy using recovered SNR that is defined as follows: given a signal  $\mathbf{x}$  and its estimate  $\hat{\mathbf{x}}$ ,  $\text{SNR} = 20 \log_{10}(\|\mathbf{x}\|/\|\mathbf{x} - \hat{\mathbf{x}}\|)$ .

**Images.** We learn a two-scale dictionary on image patches of dimension  $8 \times 8$ . Each patch was scaled down to  $4 \times 4$  size. The low resolution dictionary was formed of 64 atoms and the high resolution dictionary had 1024 atoms. We

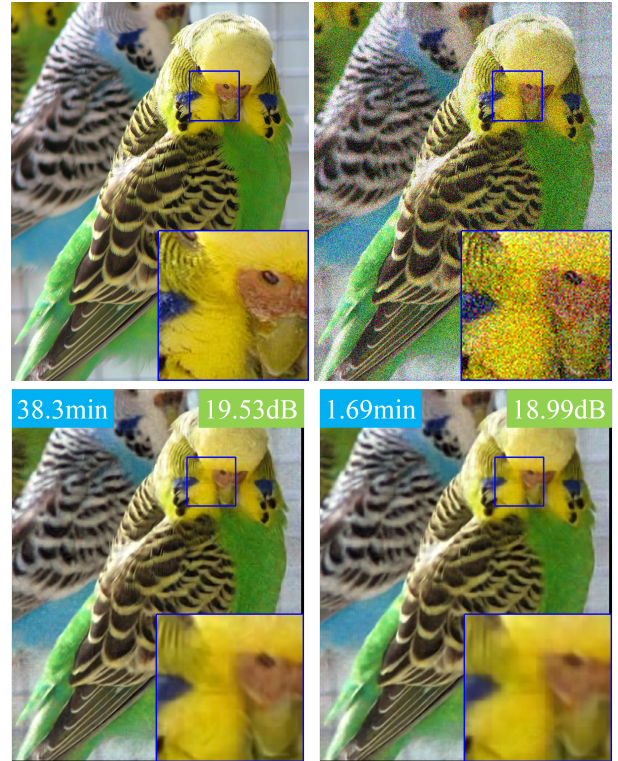


Figure 5: Visualization of results for image denoising. Clockwise from top left, original image, noisy image with SNR of 10dB, recovered image using proposed method, and recovered image using K-SVD learned dictionary. We obtain a speedup of  $22\times$  with hardly any reduction in accuracy.

compared the results against a 1024 atom single scale dictionary trained using K-SVD algorithm. We performed image denoising with the learned dictionaries. Figure 9(a), (d) show performance metrics in terms of recovered SNR for

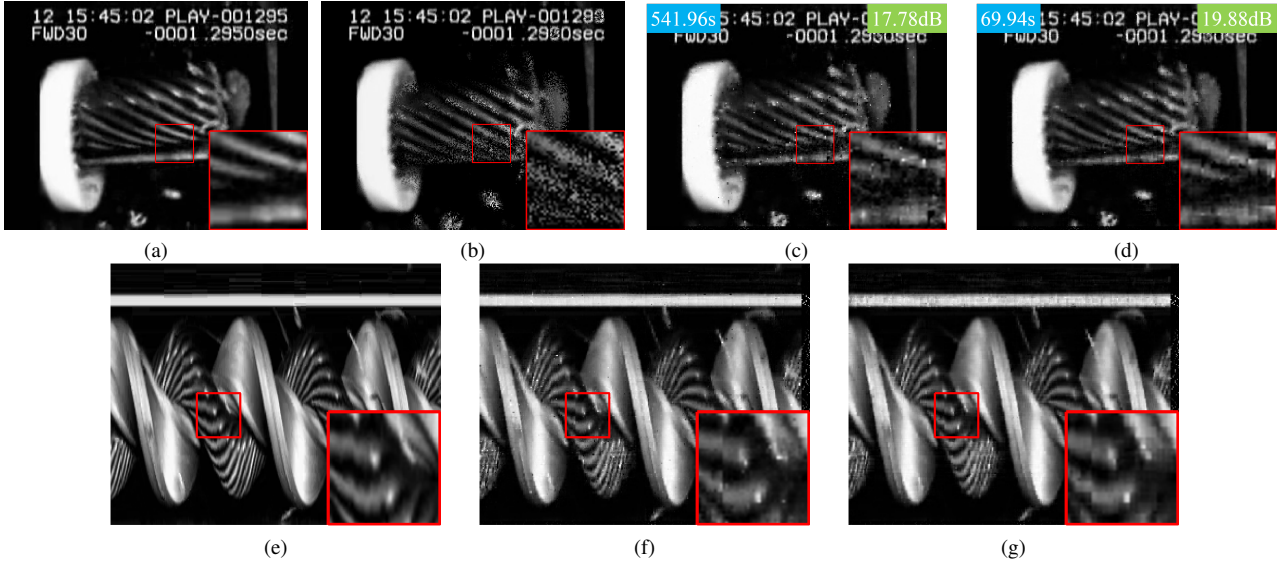


Figure 6: Visualization of video compressive sensing. We simulated the architecture proposed in Hitomi et al. [16] where a single coded image is obtained by temporal sampling of 8 frames. Given this single coded image, we recover 8 frames (at  $8\times$  frame-rate) by solving an inverse problem. (a) Ground truth video frame, (b) Coded image from 8 frames; frame recovered using (c) K-SVD dictionary and (d) proposed method; (e) XT slice for ground truth video; XT slice for video recovered using (f) single scale dictionary and (g) proposed method. Our algorithm offers significant speed ups with little loss in accuracy.

denoising and inpainting, as compared against traditional OMP. Note that speed ups in the order of  $10\times$  are obtained even for a small problem setting.

Figure 1 shows demosaicing of the Bayer pattern using both the methods. We trained an 8192 atom high resolution dictionary on  $24\times 24$  Kodak True color RGB images[1] and 512 atom low resolution dictionary on the patches downsampled to  $12\times 12$ . We compare this against 8192 atom single scale dictionary. It took 16 minutes for the single scale, whereas only 1.5 minutes for the two scale dictionary.

Figure 5 shows image denoising at an SNR of 10dB. We perform denoising with the trained RGB dictionaries of  $24\times 24$  patch and with a patch overlap of 18 pixels. With hardly any reduction in accuracy, our method performs  $22\times$  faster.

**Videos.** We trained an 8192 atom high resolution dictionary for  $8\times 8\times 16$  video patches and 512 atom low resolution dictionary for the patches downsampled to  $4\times 4\times 8$ . We compared the trained dictionaries against an 8192 atom single scale dictionary obtained using K-SVD. We maintained the same sparsity across all the dictionaries. Figure 9(b), (d) show the performance of our proposed method and conventional K-SVD+OMP for denoising and video compressive sensing where we implemented the temporal sampling method proposed in Hitomi et al. [16]. Visualization of the recovered frames is shown in Figure 6. The reconstruction accuracy is similar for both methods.

**Hyperspectral images.** We trained over-complete dictionaries from 32 channel (31 channels + 1 channel repeated for computational ease) hyper-spectral images obtained from [9]. An 8192 atom high-resolution dictionary for  $6\times 6\times 32$  patch and 512 atom low resolution dictionary for the downsampled  $3\times 3\times 8$  was trained using our proposed method, which was compared against an 8192 atom dictionary learnt using K-SVD. We tested the dictionaries for denoising and image demosaicing, the results of which are shown in Figure 9(c), and (f) respectively. A visualization of the demosaiced images can be seen in Figure 7.

**Light-field images.** We trained over-complete dictionary from data obtained from [26]. A 32768 atom high-resolution dictionary for  $4\times 4\times 8\times 8$  patch and a 1024 atom low resolution dictionary for uniformed downsampled  $2\times 2\times 4\times 4$  patch was trained using the proposed method, which we compare against a 32768 atom dictionary learned using K-SVD. We tested the dictionaries for denoising and image reconstruction using compressive angle sampling of light-field data [23], results of which are shown in Figures 9(d), and (h) respectively. Reconstruction from compressively sampled light-field on synthetic data from [23] is shown in Figure 8. We obtained a speed up of  $15\times$  for reconstruction from compressive sampling.

**Summary.** Table 1 and Figure 9 quantify the performance of the proposed signal model and those obtained using K-SVD for a wide range of parameters as well as signals.



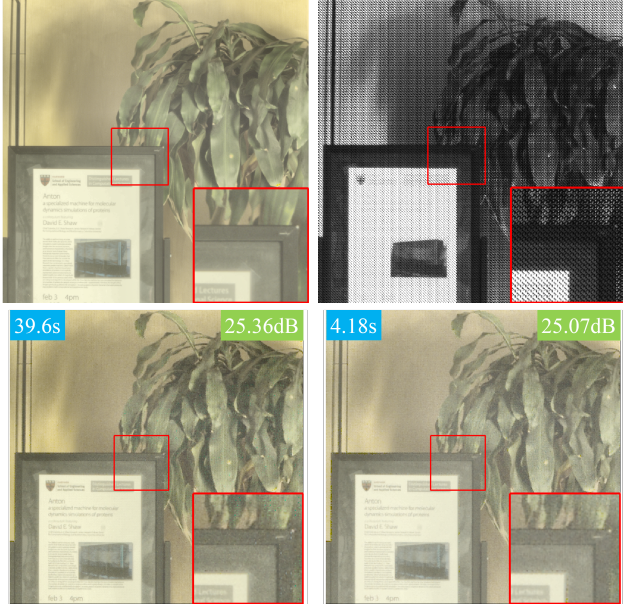


Figure 7: Demosaicing of hyperspectral images. Clockwise from top left: Original image; image obtained with randomly sampling the channels, i.e.  $I_{coded}(x, y) = I(x, y, c)$ , where  $c$  is randomly chosen from 0 to 31; image recovered using multi-scale dictionary and image recovered using single scale dictionary.

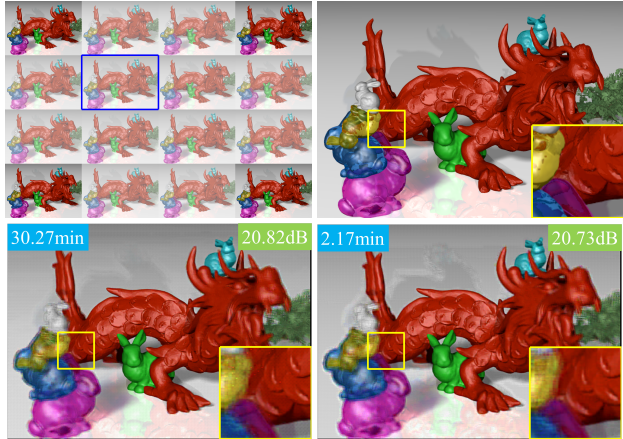


Figure 8: Compressive light-field sensing; clockwise from top left: The sixteen sub-aperture images with the high-lighted images used for reconstructing all the sixteen images; One of the unknown sub-aperture image highlighted in blue box in the previous image; The sub-aperture image recovered using multi-scale dictionary; image recovered using single scale dictionary.

Across the board, we observe that the proposed framework provides approximations that are as good as those obtained with K-SVD, but with speedups that are  $4 - 10\times$  for small-

sized problems and  $20 - 110\times$  for larger problems. The speedups obtained are comparable to results in [4] with higher approximation accuracies for our proposed method.

As a result of speed up of the sparse coding step, we get significant speed ups during the training phase ( $2 - 40\times$ ) using modified K-SVD, which makes it feasible to deal with very large scale problems.

## 5. Conclusion and discussions

We presented a signal model that enables the cross scale predictability for visual signals. Our method is particularly appealing because of the simple extension to the existing OMP and K-SVD algorithms while providing significant speed ups at little or no loss in accuracy. The computational gains provided by our algorithm are especially significant for problems involving high-dimensional dictionaries with a large number of atoms.

**Beyond two scales.** All our experiments are in the setting of two-scale dictionaries. Extending them to more scales will give significant speedups for very large dictionaries for high-dimensional problems. However, with increasing problem size, the size of training dataset also grows significantly and can potentially become a bottleneck towards the training of stable dictionaries.

**Connections to super resolution using sparse representations.** Roman et al. [33] learn a pair of low resolution and high resolution dictionary using the same sparsity pattern for the two dictionaries. Given a low resolution patch  $y_{low}$ , they solve the sparse approximation problem  $y_{low} \approx D_{low}s$  and then super resolve the image as  $y = D_{high}s$ . In contrast, our method requires the high resolution image, and uses the sparse representation of the downsampled image to predict the high resolution sparse representation. While the primary aim of [33] is image-based super resolution, our method can accommodate any inverse problem based on sparse approximation.

## 6. Acknowledgment

The authors gratefully acknowledge support from Intel Corporation.

Signal Class	$N$	$N_{low}$	$T_{low}$	$T_{high}$	$K_{low}$	$K_{high}$	Speedup	Model Accuracy (dB)	K-SVD Accuracy (dB)	Legend
Images	8x8	4x4	64	1024	8	8	4.10	20.67	21.98	$N$ Size of the high resolution signal
	24x24x3	12x12x3	512	8192	8	8	22.6	19.64	20.57	$N_{low}$ Size of the low resolution signal
Videos	8x8x16	4x4x8	512	8192	16	16	15.87	22.62	24.09	$T_{low}$ Number of atoms in low resolution dictionary
	8x8x16	4x4x8	512	8192	14	16	15.80	22.75	24.09	$T_{high}$ Number of atoms in high resolution dictionary
	8x8x32	4x4x16	512	8192	16	16	23.81	20.72	21.36	$K_{low}$ Sparsity used in low resolution dictionary
	8x8x16	4x4x8	512	16384	16	16	16.89	21.84	23.27	$K_{high}$ Sparsity used in high resolution dictionary
Hyperspectral Images	4x4x32	2x2x8	256	8192	6	6	21.19	23.90	25.62	$Speed\ up$ Ratio of time taken for single scale approximation by time taken for two scale approximation
	6x6x32	3x3x8	512	8192	8	8	29.34	22.79	25.85	
	6x6x32	3x3x4	512	8192	8	8	31.93	22.83	25.85	
Light-field Images	8x8x32	4x4x8	512	8192	8	8	27.58	23.25	24.59	
	4x4x8x8	2x2x4x4	2048	32768	4	4	111.02	19.91	21.73	
	4x4x8x8	2x2x4x4	2048	32768	6	6	108.3	22.94	23.45	

Table 1: Table with speed up for various dictionary sizes, patch sizes and sparsity. The speed up shown are for solving sparse approximation problems and quantify the ratio of time taken by OMP using a K-SVD learnt dictionary to zero tree OMP on the proposed model. Also shown are approximation errors on training dataset for both K-SVD and the proposed algorithm.

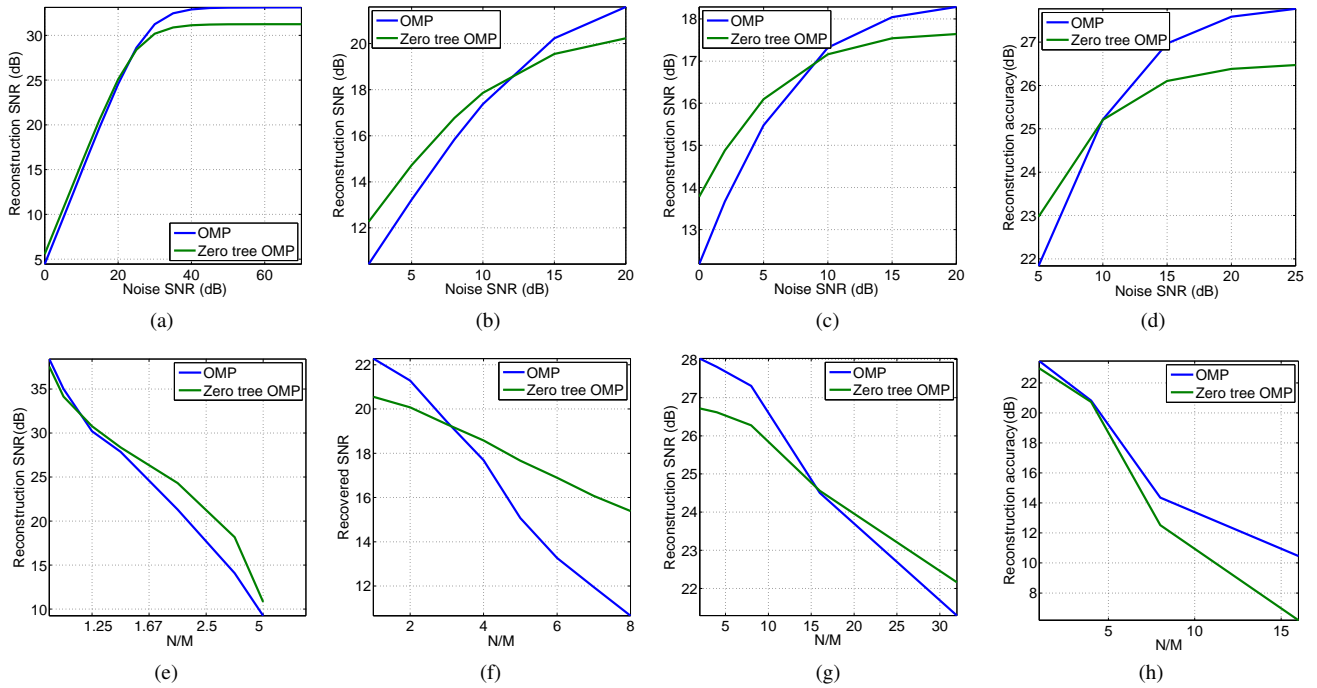


Figure 9: Comparison of zero tree OMP vs. OMP based processing for various applications. (a) Image denoising, (b) Video denoising, (c) Hyperspectral image denoising, (d) Light-field denoising, (e) Image inpainting ( $N/M$  is the number of unknown pixel values per each known), (f) Video compressive sensing using coded images ( $N/M$  represents the number of frames recovered from each coded image), (g) Hyperspectral image demosaicing ( $N/M$  is the number of spectral channels combined into one image), and (h) Light-field compressive sensing using random angle sampling ( $N/M$  is the number of sub-aperture images reconstructed from a sub-aperture image). Observe that the curves for OMP as well as our proposed algorithm are very comparable.



## References

- [1] Kodak lossless true color image suite. <http://r0k.us/graphics/kodak/>. Accessed: 2015-10-05. 6
- [2] E. Adelson, E. Simoncelli, and W. T. Freeman. Pyramids and multiscale representations. *Representations and Vision*, Gorea A.,(Ed.). Cambridge University Press, Cambridge, pages 3–16, 1991. 1
- [3] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Processing*, 54(11):4311–4322, 2006. 2
- [4] A. Ayremlou, T. Goldstein, A. Veeraraghavan, and R. G. Baraniuk. Fast sublinear sparse representation using shallow tree matching pursuit. *arXiv preprint arXiv:1412.0680*, 2014. 2, 7
- [5] R. G. Baraniuk. Optimal tree approximation with wavelets. In *SPIE Intl. Symp. Optical Science, Engineering, and Instrumentation*, 1999. 1, 3
- [6] R. G. Baraniuk. Compressive sensing. *IEEE Signal Processing Magazine*, 24(4), 2007. 3
- [7] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Trans. Information Theory*, 56(4):1982–2001, 2010. 1
- [8] J. Bieniarz, R. Muller, X. Zhu, and P. Reinartz. On the use of overcomplete dictionaries for spectral unmixing. In *4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2012. 3
- [9] A. Chakrabarti and T. Zickler. Statistics of Real-World Hyperspectral Images. In *IEEE Conf., Computer Vision and Pattern Recognition*, 2011. 6
- [10] D. G. Dansereau, O. Pizarro, and S. B. Williams. Decoding, calibration and rectification for lenselet-based plenoptic cameras. In *IEEE Conf., Computer Vision and Pattern Recognition*, 2013.
- [11] S. Deutsch, A. Averbush, and S. Dekel. Adaptive compressed image sensing based on wavelet modeling and direct sampling. In *SAMPTA*, 2009. 1
- [12] M. Elad. *Sparse and redundant representations: From theory to applications in signal and image processing*. Springer, 2010. 2
- [13] Y. Fang, L. Chen, J. Wu, and B. Huang. GPU implementation of orthogonal matching pursuit for compressive sensing. In *IEEE Intl. Conf on Parallel and Distributed Systems (ICPADS)*, 2011. 2
- [14] M. Gharavi-Alkhansari and T. S. Huang. A fast orthogonal matching pursuit algorithm. In *IEEE Intl. Conf. Acoustics, Speech, Signal Processing*, 1998. 2
- [15] R. Gribonval. Fast matching pursuit with a multiscale dictionary of gaussian chirps. *IEEE Trans. Signal Processing*, 49(5), 2001. 2
- [16] Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, and S. K. Nayar. Video from a single coded exposure photograph using a learned over-complete dictionary. In *IEEE Intl. Conf. Computer Vision*, 2011. 2, 3, 6
- [17] R. Jenatton, J. Mairal, F. R. Bach, and G. R. Obozinski. Proximal methods for sparse hierarchical dictionary learning. In *Intl. Conf., Machine Learning*, 2010. 2
- [18] C. La and M. N. Do. Tree-based orthogonal matching pursuit algorithm for signal reconstruction. In *IEEE Intl. Conf. Image Processing*, 2006. 2
- [19] S. Li and H. Qi. Sparse representation based band selection for hyperspectral images. In *IEEE Intl. Conf. Image Processing*, 2011. 3
- [20] X. Lin, Y. Liu, J. Wu, and Q. Dai. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Trans. Graphics*, 3
- [21] B. Mailhé, R. Gribonval, F. Bimbot, and P. Vandergheynst. A low complexity orthogonal matching pursuit for sparse signal approximation with shift-invariant dictionaries. In *IEEE Intl. Conf. Acoustics, Speech, Signal Processing*, 2009. 4
- [22] J. Mairal, G. Sapiro, and M. Elad. Multiscale sparse image representation with learned dictionaries. In *IEEE Intl. Conf. Image Processing*, 2007. 2
- [23] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar. Compressive light field photography using overcomplete dictionaries and optimized projections. *ACM Trans. Graphics*, 32:46, 2013. 3, 6
- [24] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision research*, 37(23):3311–3325, 1997. 1, 2
- [25] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Asilomar Conf. Signals, Systems, Computers*, 1993. 2
- [26] S. Pelletier. *Acceleration methods for image super-resolution*. PhD thesis, McGill University, 2009. 6
- [27] A. Secker and D. Taubman. Lifting-based invertible motion adaptive transform (limat) framework for highly scalable video compression. *IEEE Trans. Image Processing*, 12(12):1530–1542, 2003. 1
- [28] J. M. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Trans. Signal Processing*, 41(12):3445–3462, 1993. 1, 3
- [29] S. Tambe, A. Veeraraghavan, and A. Agrawal. Towards motion aware light field video for dynamic scenes. In *IEEE Intl. Conf. Computer Vision*, 2013. 3
- [30] J. J. Thiagarajan, K. N. Ramamurthy, and A. Spanias. Learning stable multilevel dictionaries for sparse representations. *IEEE Trans. Neural Networks and Learning Systems*, PP(99), 2014. 2
- [31] S. N. Vitaladevuni, P. Natarajan, and R. Prasad. Efficient orthogonal matching pursuit using sparse random projections for scene and video classification. In *IEEE Intl. Conf. Computer Vision*, 2011. 2
- [32] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky. Random cascades on wavelet trees and their use in analyzing and modeling natural images. *Appl. Comp. Harmonic Analysis*, 11(1):89–123, 2001. 1
- [33] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces*, pages 711–730. Springer, 2012. 7